

사용자 이동 방식 인지를 위한 위치정보 수집 및 분석

윤용상, 김경백
전남대학교 전자컴퓨터공학부
e-mail : tiah.florens@gmail.com, kyungbaekkim@jnu.ac.kr

Collection and Analysis of Location Data for Recognizing User Movement Methods

Yongsang Yoon, Kyungbaek Kim
Dept. of Electronics and Computer Engineering,
Chonnam National University

요 약

최근 모바일 기기의 고도화 및 위성 기술의 발달에 따라, 모바일 단말에서 개인 위치의 수집이 용이해지고 있고, 이에 따라 개인 위치 정보 기반의 다양한 서비스들이 주목 받고 있다. 수집된 위치 정보를 필요한 기준에 따라 적절하게 분석한다면 다양한 위치기반 서비스 및 개인용 스마트 기기 인터페이스 등을 위한 매우 유용한 정보로 활용 할 수 있다. 예를 들어 각 지역별 유동인구 또는 사람들이 밀집한 특정 지역이나, 시간대를 확인 하여 위치기반 서비스의 성능을 향상 시킬 수 있다. 또한 스마트 기기에서 사용자의 위치와 연관된 이동 방식을 인지하여 개인 사용자에게 필요한 인터페이스를 제공할 수 있다. 이 논문에서는 위치 정보 수집을 위한 틀에 대한 설명과, 약 1개월간 수집된 위치정보를 기반으로 분석 결과를 소개한다. 이 결과를 토대로 이동 방식 인식을 위한 알고리즘 개발 시 필요한 점들을 고찰한다.

1. 서론

불과 몇 년 전까지만 하더라도 핸드폰의 주된 목적은 전화나 문자였다. 하지만 오늘날의 핸드폰은 스마트폰이라 불리며 대다수 사람들의 필수품이 되었다. 사용자들은 스마트폰으로 이메일을 확인하거나, youtube 감상 혹은 SNS를 이용하기도 하고, 멀티미디어 콘텐츠 등을 즐긴다. 이 과정에서 다양한 종류의 데이터가 스마트폰에 발생한다. 즉 스마트폰에는 개개인의 대한 여러 가지 정보가 저장되어 있으며, 이를 분석함으로써 개개인의 사용자에게는 더욱 편리한 서비스를 제공할 수 있고, 전체 시스템을 보다 효율적으로 관리 할 수 있다. 이 논문에서는 이러한 개인 정보 데이터들 중 하나인 위치정보에 주목하였다.

위치 정보를 적절히 분석한다면 매우 유용한 정보가 될 것으로 기대된다. 예를 들어 기업은 시간대, 연령대, 성별을 기준으로 사람들이 많이 몰리는 곳에 알맞은 상품을 판매하거나, 지점을 효율적으로 관리 할 수 있다. Google 혹은 Apple 과 같은 기업들은 자사 회원의 위치 정보를 모아 분석 한 후 자주 방문하는 지역, 또는 특정 지역에서 많이 검색되는 특정 단어를 추출해 개인별 맞춤 광고 서비스를 제공한다. 국가에서는 자원을 적절한 곳에 투자하

거나 공공시설을 늘려 시민들의 편의성을 증대 시킬 수 있다. 개인적으로는 자신의 기록을 확인하거나 핸드폰 분실 시 위치 기록을 추적할 수도 있을 것이다. 또한, 개인의 위치정보를 기반으로 개인의 행동 양식을 분석 할 수 있다면 상황에 알맞은 스마트폰 서비스를 제공할 수 있다.

이 논문에서는 사용자의 위치정보 분석을 통한 다양한 서비스 제공을 위한 초기 단계인 위치정보 수집 및 분석 틀에 대한 개발 및 약 1개월간 수집된 위치정보를 기반으로 한 테스트 결과에 대해 설명한다. 사용자 행동 양식중 하나인 사용자 이동 방식 인식을 위한 분석결과를 설명한다.

2. 위치정보 수집 및 분석 틀

사용자 위치정보 수집을 위해 우리는 최근 Android OS 기반 스마트폰에 기본으로 지원되는 GPS를 사용한다. 개발된 틀은 일정 주기마다 스마트폰 내의 Tiny DB에 임시로 저장되고, 저장된 정보는 파일 또는 Internet을 통해 Main DB로 추출된다. 이렇게 추출된 위치정보 데이터는 Google MAP과 연동하여 분석된다.

2.1 수집 기능 및 분석 옵션

사용자 위치정보 수집은 Android에서 제공하는 Location Provider의 GPS를 사용하였다. 이때, 주기와 Threshold

[†] “본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터육성 지원사업의 연구결과로 수행되었음” (NIPA-2013-H0301-13-3005)

Distance를 설정할 수 있게 하여 모바일 기기의 배터리 수명 및 사용자의 기호에 따라 위치정보 수집을 실행 할 수 있도록 하였다. 수집된 위치정보를 저장할 때는 각 사용자의 정보(ID 및 사용자 그룹), 위치정보(위도, 경도), 시간 정보(시간 및 날짜) 를 하나의 튜플로 작성하여 저장 하였다.



(그림 1) 위치정보 분석 툴 기본화면

수집된 데이터의 분석을 위해 그림 1과 같이 다양한 입력 파라미터를 가지는 Google Map기반 분석 툴을 개발 하였다. 입력 파라미터로는, Data 생성 시간 및 날짜, Data 대상 그룹 및 ID, 그리드 크기, 데이터 분석 타입 등이 있다. Data생성 시간 및 날짜는 분석 하고자하는 위치 정보의 기간 설정을 위해 사용된다. Data대상 그룹 및 ID 옵션은 분석하고자 하는 대상이 특정 개인인지 아니면 특정 그룹인지를 설정 할 수 있게 도와준다. 즉 전남대학교 내에서 학생과 교직원간의 그룹핑, 각 학과 학생별 그룹핑, 학년별 그룹핑 등을 생각 할 수 있다. 그리드 크기 옵션은 분석할 위치정보의 입상도(granularity)를 나타낸다. 그리드의 크기를 줄일수록 보다 자세한 위치정보의 분석이 가능해지는 반면 계산 소요시간이 늘어나게 된다. 반면 그리드의 크기가 너무 커지면 정확한 분석이 힘들 수 있게 된다. 데이터 분석 타입 옵션은 각 그리드에서 위치정보의 개수를 셀 때, 사용자 기반으로 할 지 아니면 정보 기반으로 할지를 설정한다. 즉 특정지역을 지나치는 사람이 있을 때, 이 사람의 속도에 따라 해당 그리드에 저장되는 위치정보의 개수가 달라지게 된다. 이 이유 때문에 임의의 그리드에서 소수의 사람이 다수의 위치정보를 제공하는 반면 또 다른 그리드에서는 다수의 사람이 소수의 위치정보를 제공할 수 있게 된다. 이러한 점들을 분석하기 위해 데이터 분석 타입 옵션을 GPS LOG기반과 INDIVIDUAL PERSON기반의 두 가지로 제공 한다.

이와 같이 입력된 파라미터들을 기반으로 Map상의 각 그리드에서 표현할 위치정보의 개수를 구하고, 이 개수의 정도에 따라 그리드의 불투명도를 조절한다. 현재 개발된 툴에서는 위치정보의 개수가 많을 수록 불투명도가 높아 지도록 하였다.

2.2 위치정보 수집 및 분석 툴 테스트 결과

개발된 위치정보 수집 및 분석 툴의 테스트를 위해 약 1개월간 4명의 전남대학교 공대 학생으로부터 위치정보 로 그를 수집하였다. 이에 대한 분석 결과를 그림 2와 그림 3과 같다.

그림 2는 한 달 동안의 기간 중 오후 5시부터 밤 10시 사이의 위치정보를 GPS LOG기반으로 분석할 결과이다. 이 그림에서 학교 쪽의 불투명도가 가장 높고, 광주 전역의 버스 노선을 중심으로 학생들의 위치가 확인된 그리드들이 불투명하게 처리 되는 것이 관찰 되었다. 이는 귀가 시간 직전 학생들이 학교 쪽에 밀집해 있어 학교 쪽의 불투명도가 높은 것으로 분석되고, 이후 학생들이 흩어져서 귀가함에 따라 버스노선상의 그리드의 불투명도는 상대적으로 낮은 것으로 분석 된다.



(그림 2) 위치정보 분석결과: 오후 5시 ~ 오후 10시, Large Grid Size, GPS COUNT기반 분석



(그림 3) 위치정보 분석결과: Large Grid Size, INDIVIDUAL PERSON기반 분석

그림 3에서는 INDIVIDUAL PERSON기반으로 분석된 결과 중 전남대 주변을 나타낸 것으로, 대부분의 데이터가 공과 대학 쪽에 밀집되어 있음을 알 수 있다. 이는 위치정

보 수집에 참여한 학생들이 모두 공과대학 학생이라는 점이 영향을 미친 것으로 분석 된다.

3. 사용자 이동방식 패턴 분석

이상의 개발된 틀은 여러 사용자들의 위치정보를 전체적으로 분석하여 다양한 위치기반 서비스를 위한 입력 정보를 제공한다. 이에 덧붙여 위치정보를 기반으로 개인의 행동양식을 분석할 수 있다면 각 개인에 적합한 서비스를 제공할 수 있을 것이다. 이러한 개인의 행동양식 분석을 위한 초기 단계로 사용자 이동 방식을 분석해 보았다.

3.1 이동 벡터 정보 추출

이를 위해 우선 각 개인의 위치정보를 이동 벡터 정보로 변환하는 것이 필요하다. 현재 저장되어 있는 위치정보 튜플은 위치정보 데이터 ID, 사용자 ID, 사용자그룹, 위도, 경도, 타임스탬프 (d_id, u_id, u_g, lat, ngt, ts)로 이루어져 있다. 그리고 사용자 이동 벡터 정보 튜플은 벡터 정보 데이터 ID, 사용자 ID, 사용자 그룹, 벡터 시작 위치 데이터 ID, 벡터 끝 위치 데이터 ID, 벡터의 길이(거리, m), 벡터의 시간차(sec), 이동 속도(m/s) (v_id, u_id, u_g, vs_d, ve_d, dist, t, v) 로 이루어져 있다. 즉 위치 정보 중 두 개의 연속한 사용자 위치정보간의 거리, 시간차 그리고 이동 속도를 계산하여 벡터 정보로 변환 한다.

두 위치 정보의 시간차, t,는 저장되어 있는 타임스탬프를 비교함으로써 쉽게 계산할 수 있다. 두 위치 정보의 거리를 구하기 위해서, 우리는 Spherical law of cosine을 사용하였다. 이 법칙은 임의의 Sphere표면상의 두 위치들에 대한 radian표현법을 알고 있다면, 두 위치사이의 arc의 길이를 구하는 법을 정의한다. 즉, $dist = \arccos(\sin(\phi_1) \times \sin(\phi_2) + \cos(\phi_1) \times \cos(\phi_2) \times \cos(\Delta\lambda)) \times R$ 과 같은 식을 적용하여 거리를 구한다. 이식을 지구에 적용할 경우, ϕ 는 위도를 λ 는 경도($\Delta\lambda$ 은 두 경도의 차이)를 나타내고, R은 지구의 반지름의 길이를 의미한다. 이 식을 64비트기반의 소수점 15자리까지 표현 가능한 플로팅 포인트 수를 사용하여 계산 할 경우, 약 1m 이하의 오차를 보이게 된다.

3.3 이동 벡터 정보 분할 및 이동 방식 특성 분석

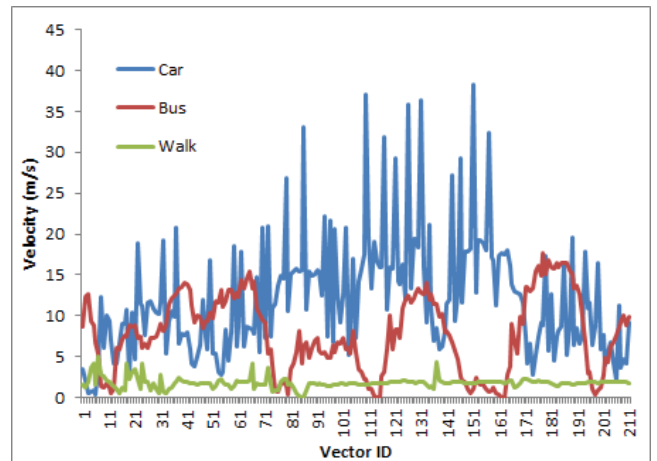
변환 작업을 통해 생성된 이동 벡터 정보는 사용자의 이동 방식 분석에 사용 된다. 이를 위해 우선 이동 벡터 정보를 의미 있는 세그먼트(segment)들로 분리 한다. 즉 벡터 정보의 시간차가 너무 큰(6000 sec이상) 정보를 기준으로 벡터의 정보를 여러 개의 세그먼트로 분리 하였다.

각기 분리된 세그먼트들은 사용자의 입력 시 수집된 이동 방식 정보와 매칭 시켰다. 현재 고려한 사용자 이동 방식은 도보, 버스, 자동차의 세 가지이다. 그림 4에서는 각 3가지 이동 방식과 매칭 된 이동 벡터 정보의 속도를 출력한 그래프이다. 이 그래프에서 알 수 있듯이 도보는 자동차와 버스와 속도 측면에서 확연히 차이가 남을 알 수

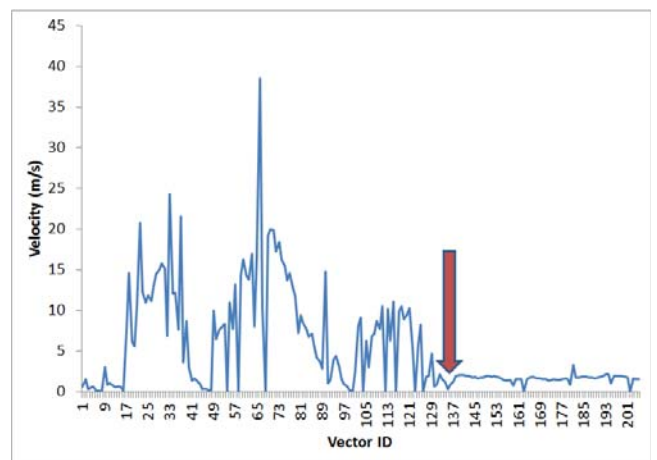
있다. 도보의 경우, 평균적으로 1.5m/s의 이동속도를 보인다. 또한 특별히 이동이 멈추는 패턴은 관찰 되지 않는다. 반면 자동차나 버스의 경우 이동 속도는 5m/에서 30m/s 사이에서 지속적으로 변경되는 것이 관찰 되었다.

자동차와 버스의 이동 벡터 정보에서의 다른 점은 크게 두 가지이다. 첫 번째는 Stopping Rate이다. 즉 이동벡터의 속도가 0에 근접할 정도로 떨어지는 Stopping구간이 주기적으로, 그리고 보다 자주 발생할 경우 이 이동 벡터 정보는 버스의 이동 벡터일 확률이 높다는 것이다. 두 번째는 가속 정도이다. 자동차의 경우 이동시 순간 가속이 버스보다 크기 때문에 자동차와 매칭이 된 이동 벡터 정보에서는 아주 큰 속도를 가지는 벡터 정보를 관찰 할 수 있다.

이러한 고려점들은 임의의 이동 벡터 세그먼트와 매칭이 되는 사용자 이동 방식을 찾기 위한 디지전 트리 기반의 알고리즘에 적용 될 수 있다.



(그림 4) 사용자 이동 방식에 따른 이동 벡터 정보



(그림 5) 버스와 도보 정보로 이루어진 하나의 이동 벡터 세그먼트

이러한 디지전 알고리즘을 적용하기 이전에 세그먼트를 분리하는 과정에서 고려할 점을 발견 할 수 있었다. 그림 5는 시간차를 기준으로 이동 벡터 정보를 다수의 세그먼트로 나눈 후 얻어진 임의의 세그먼트의 벡터 정보를 보여주고 있다. 이 세그먼트는 사용자가 버스를 타고 이동하던 도중 버스에서 내린 후, 계속해서 도보로 이동을 계속하는 상황과 매칭이 되어 있다. 그림 5의 빨간색 화살표가 가리키는 부분은 사용자가 버스에서 내리는 시점이다. 이처럼, 사용자의 이동 방식은 시간에 따라 연속적으로 변할 수 있기 때문에 단순한 시간 기반의 세그멘테이션을 통해서 인지하기 힘들 수 있다는 점을 알 수 있었다.

4. 결론

주목 받고 있는 위치정보를 적절한 기준으로 분석함으로써 목적에 따라 유용한 정보로 활용 할 수 있게끔 하는 분석 툴을 개발하고자 했다. 약 1개월 간의 데이터 수집 과 초기 분석을 통해, 개발된 툴의 기능을 확인하였고, 향후 많은 데이터의 확보를 통해 보다 의미 있는 분석 결과를 얻을 수 있을 것으로 예상된다. 또한 이미 알고 있는 사용자의 위치정보와 이동 방식 간의 매칭을 분석하여 사용자의 이동 방식 인식을 위한 세그멘테이션 및 디지전 알고리즘을 개발할 때 필요한 고려점들을 살펴보았다. 향후, 사용자 이동 방식 인식을 위한 세부 알고리즘을 개발하고 이를 응용할 수 있는 위치 기반 시스템에 대한 연구를 지속할 계획이다. 또한 지리/지역 정보와 이동 벡터 정보의 매칭을 통해 보다 정확한 사용자 행동 패턴을 인식할 수 있는 기법을 연구할 계획이다.

참고문헌

- [1] Roger S. Pressman "Software Engineering A Practitiners' Approach" 3rd Ed. McGraw Hill
- [1] MySQL Tutorial : <http://dev.mysql.com/doc/refman/5.0/en/tutorial.html>
- [2] Google Maps JavaScript API V3 : <https://developers.google.com/maps/documentation/javascript/?hl=ko>
- [3] Professional Android Application Development , Wrox , Reto Meier
- [4] JSP 2.1 웹프로그래밍 , 최범균 , 가메출판사
- [5] UnderStanding Mobility Based On GPS Data , Yu Zheng, Quannan Li , Yukun Chen ,Xing Xie , Wi Ying Ma, In Proceedings of UbiComp 2008.
- [6] Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web, Yu Zheng, Like Liu, Longhao Wang, Xing Xie, In Proceedings of WWW 2008.